Method for labeling solid, liquid and gaseous substances

The invention relates to a method for labeling solid, liquid and gaseous substances.

According to the prior art, various methods are known, which utilize nucleic acid sequences for labeling.

WO 90/14441 discloses a method in which a predefined specific nucleic acid sequence is amplified and subsequently identified.- This cannot be used for complex labeling in the style of a code.

WO 91/17265 describes a method for labeling a material with microtraces of DNA. The DNA which is present in the smallest amounts is amplified by a PCR. The DNA is identified by sequencing. Detection of the DNA employed for labeling is time-consuming and complicated.

US 5,866,336 discloses methods in which a formation of intra- or intermolecular hybrids is detected by energy transfer between donor and acceptor molecules.

WO 94/04918 utilizes a nucleic acid bound to a particle for labeling a liquid. The nucleic acid is amplified and identified, for example, by means of a radioactive label.

WO 95/02702 discloses a method which uses various nucleic acids in combination with various particles. This can be used to generate a complex labeling code. The nucleic acids are identified by means of PCR amplification, where appropriate in combination with sequencing. In order to identify the code, furthermore identification of the particles and combination with the information obtained from the identifications are required.

US 5,656,731 describes a labeling of antibodies by means of nucleic acids. The presence of the sought-after antibody is carried out [sic], after a preceding selection process, on the basis of detecting the nucleic acid. For this purpose, the nucleic acid is amplified.

US 5,708,153 describes a labeling of polymeric substances. In this connection, simultaneously with the synthesis of the polymeric substance, a nucleic acid monitoring the synthesis is synthesized. The composition of the polymeric substance is identified by amplifying and sequencing the nucleic acid.

The methods known are costly and time-consuming. As far as a sequencing reaction is required for identifying the label, utilization of a complex labeling code or identification of a label from a plurality of labels is possible only with extreme difficulties.

It is an object of the invention to indicate a method which can be used to label unambiguously a plurality of substances and to identify them rapidly and inexpensively.

This object is achieved by the features of claim 1. Expedient embodiments result from the features of claims 2 to 21.

In accordance with the invention a method for labeling and identifying solid, liquid and gaseous substances is provided for wherein labeling is carried out by selecting at least one nucleic acid sequence from a first group of predefined nucleic acid sequences having in each case a labeling [sic] sequence section and adding it to the substance,

wherein a second group of further nucleic acid sequences which have in each case a detection sequence section complementary to one of the identification sequence sections is provided for identification,

wherein first melting points of hybrids formed from the identification sequence sections together with the detection sequence sections complementary thereto differ by not more than 5°C from one another and

second melting points of not completely complementary hybrids from the identification sequence sections and detection sequence sections are more than 5°C lower than the lowest of the first melting points and

wherein identification is carried out by contacting the nucleic acid sequence(s) selected from the first group with the further nucleic acid sequences of the second group under predefined hybridization conditions and detecting hybridization.

The method can be carried out easily, rapidly and inexpensively. It is possible to amplify and identify all identification sections used for labeling in a single reaction mixture.

Nucleic acid sequences here mean both single- and double-stranded sequences which essentially consist of nucleic acids.

According to one embodiment feature, the identification sequence section is located between two primer binding sequence sections. - In this case, the nucleic acid sequence may be single-stranded. Amplification is readily possible, for example, by means of the polymerase chain reaction (PCR).

According to a further embodiment, in each case two nucleic acid sequences have a part section of a common

identification sequence section at their 5' end and a primer binding sequence section is bound to said part section. In this embodiment the identification sequence section is initially not present but is generated only during the amplification reaction. Advantageously, the part sections are in this connection partly complementary to one another. - This increases the reliability of the labeling.

Expediently, the primer binding sequence sections have the same melting point. This makes possible a simultaneous amplification reaction of the nucleic acid sequences in a single reaction mixture.

The nucleic acid sequences may be amplified preferably by means of PCR and by using fluorescent primers. In order to improve the stability of the label, it is advantageously provided for the nucleic acid sequences to be linked on at least one end to an agent which counteracts degradation caused by exonuclease.

Complex labels may be generated by providing the nucleic acid sequences with a specific coupling group. The coupling group may be selected from the following group [sic]: biotin group, amino group, thiol group or hapten. By means of the coupling group the nucleic acid sequence provided therewith can be bound specifically and/or, where appropriate, also be identified.

Further primer binding sequence sections may also be used for identification. Each primer binding sequence section can be identified unambiguously on the basis of a specific fluorophoric group bound thereto.

Identification can be made easier further by a molecule carrying a fluorophoric group, which is bound to the nucleic acid sequence.

In particular when using identical identification sequence sections, it is possible also to use further primer binding sequence sections for identification. Each of the primer binding sequence sections can be identified unambiguously on the basis of a specific fluorophoric group bound thereto.

It has proved to be expedient to bind the nucleic acid sequence to the substance and to use as substance one of the following agents: antibodies, lectins, receptors, nucleotide sequences, PNA sequences, peptides, proteins, sugars, ligands. It is regarded as particularly advantageous to bind the nucleic acid sequences to particles or to include them therein. The particles may be from 30 nm to 3 mm in size. They are advantageously silica, polystyrene, polyvinyl chloride, polyethylene, nylon or glass milk particles. However, the particle may also be a viral capsid or a virus-like particle. - The use of nucleic acid sequence-carrying particles is particularly advantageous, because it is possible, owing to the size of the particles, to sort and isolate them, for example, in a particle sorter. The particle may be bound as label carrier to the substance to be labeled.

It has proved advantageous for identification that each of the further nucleic acid sequences is bound to a predefined site of a solid surface, preferably on a chip, a microtiter plate or film. This makes it possible to decode complex labeling codes or a multiplicity of codes in a single process step.

Hybridization of an identification sequence section with a complementary detection sequence section can be detected by means of fluorescence. The proposed identification can be detected particularly easily.

According to a further embodiment, it is provided for at least two nucleic acid sequences to be added to the

substance as a label. It is thereby possible to provide a complex code by using a small number of different identification sequence sections.

The nucleic acid sequences and/or the further nucleic acid sequences are preferably prepared synthetically.

It is also possible to use chimeras of nucleic acids and nucleic acid analogs, such as PTO or PNA, instead of the nucleic acid sequences or the further nucleic acid sequences. Such chimeras have increased stability against enzymatic degradation.

On the basis of the drawing [sic], exemplary embodiments of the invention are illustrated in more detail below. The figures show:

Figs. 1a - e amplification of a first nucleic acid sequence,

Figs. 2a - e amplification of two second nucleic acid sequences,

Figs. 3a - d identification of a first nucleic acid sequence by means of molecular beacons,

Fig. 4 labeling of an identification sequence section,

Fig. 5a nucleic acid sequences labeled with fluorophoric groups,

Figs. 5b, c a detector with the further nucleic acid sequences,

Figs. 6a - d selection and identification of third nucleic acid sequences,

Figs. 7a - f preparation of a complex code,

Figs. 8a - d preparation of labeling particles,

Figs. 9a - d labeling of molecules with labeling particles according to Fig. 8 and

Fig. 10 selection and identification of substances labeled with labeling particles.

Figs. 1a - e depict diagrammatically a first nucleic acid sequence N(I)1 and amplification thereof. The first nucleic acid sequence N(I)1 has at its 3' end a first primer binding sequence section PBS1 and at its 5' end a second primer binding sequence section PBS'2 complementary to a second primer binding sequence section PBS2 (not shown here). An identification sequence section IDS is located between the first PBS1 and the second complementary primer binding sequence section PBS'2.

For amplification, the first nucleic acid sequence N(I)1 is contacted with a first primer P1 and a second primer P2. The first P1 and the second primer P2 hybridize with the first primer binding sequence section PBS1 and the second primer binding sequence section PBS2, respectively, which are complementary thereto. The first P1 and the second primer P2 are extended by polymerase; an identification sequence section IDS' complementary to the identification sequence section IDS is formed (Fig. 1c). The second primer P2 then binds to the complementary nucleic acid sequence N'(I)1 (Fig. 1d). Then a double-stranded DNA which contains the identification sequence section IDS is formed by the polymerase (Fig. 1e).

In the exemplary embodiment depicted in Figs. 2a-e the identification sequence section IDS is initially not present. A second nucleic acid sequence N(II)1 has at

its 5′ end the first primer binding sequence section PBS1. A part section IDS-A of the common identification sequence section IDS is bound thereto. A further second nucleic acid sequence N(II)2 has at its 3′ end the second primer binding sequence section PBS2. A second part section IDS-B of the identification sequence section IDS is bound thereto. The first part section IDS-A and the second part section IDS-B are in sections complementary to one another.

Figs. 2b - e depict the amplification. The first primer P1 and the second primer P2 bind to the primer sequence sections PBS1 and PBS2, respectively, which are in each case complementary thereto. The complementary sequences of the first IDS-A and of the second part section IDS-B are synthesized by polymerase (Fig. 2c). These synthesis products may hybridize in further cycles at their 3′ ends (Fig. 2d) and may be extended (Fig. 2e). The product formed is a nucleic acid sequence which has the complete identification sequence section IDS.

Advantageously, it is possible in this method variant to use nucleic acid analogs such as, for example, PNA or PTO in the region of the first PBS1 and second primer binding sequence section PBS2. Such nucleic acid analogs have increased stability against enzymatic degradation. The stability can also be further increased by coupling the 5′ end of the nucleic acid sequence to an agent which prevents 5′ exonuclease degradation. Suitable agents for this purpose are, for example, PNA or PTO.

The primer binding sequence sections PBS1, PBS2 are advantageously chosen such that the amplification reaction can be carried out within a narrow temperature range. For this purpose, the primer binding sequence sections PBS1 and PBS2 are chosen such that their melting points differ by not more than 5° Celsius from one another. In order to increase the specificity of

the identification reaction, it is advantageous that the melting point of the primer binding sequence sections PBS1 or PBS2 in a formation of not completely complementary hybrids is more than 5° Celsius below the lowest melting point of a completely complementary hybrid. This renders impossible the formation of unspecific hybrids during amplification.

The labels are identified by hybridizing the identification sequence section IDS with a detection sequence section completely complementary thereto. In order to increase the specificity of the identification reaction, it is provided for all identification sequence sections IDS together with the detection sequence sections IDP complementary thereto to have first melting points which preferably differ by not more than 5° Celsius from one another. To further increase the specificity, it is provided for to be each melting point of an incomplete hybrid with an identification sequence section IDS to be more than 5° Celsius below the lowest melting point of completely complementary hybrids.

In order to further simplify carrying out the method, it is additionally provided for the melting points of the primer binding sequence sections PBS1, PBS2 and the identification sequence sections IDS to be essentially identical.

According to another variant of the method, a first nucleic acid sequence N(I)1 is duplicated in an amplification reaction (Figs. 3a and b). The reaction mixture additionally contains molecular beacons having detection sequence sections IDP1-n complementary to the identification sequence sections. The molecular beacons MB have the shape of a hairpin loop. A fluorophoric group F11, F12, F13 and, arranged opposite, a quencher Q1, Q2, Q3 are located in the end regions of the molecular beacons.

As soon as a sufficient number of first nucleic acid sequences N(I)1-n has been prepared by amplification, the identification sequence section IDS1-n hybridizes specifically with the detection sequence section IDP1-n complementary thereto. In this connection, the spatial relationship between the quencher Q1, Q2, Q3 and the fluorophoric group F11, F12, F13 is terminated. A specific fluorescence can be detected (Fig. 3d).

Using various fluorophores F11, F12, F13 in the molecular beacons MB makes it possible to distinguish various identification sequence sections IDS1-n.

In the variant of the method depicted in Figs. 4a-c, amplification is carried out using a first primer P1 which is labeled with a fluorophoric group F11. When using nucleic acid sequences having an identical identification sequence section IDS, it is possible to distinguish said nucleic acid sequences by using different primer binding sequence sections which are labeled in each case with specific fluorophoric groups.

The identification sequence section IDS is identified by contacting the nucleic acid sequence, prepared in the amplification using the fluorophoric group F11, with a complementary detection nucleotide sequence IDP which is bound at a predefined site of a solid surface. The fluorescence which then appears at the predefined site may be recorded by means of a conventional detection apparatus (see Figs. 5a - c).

According to another variant of the method, the nucleic acid sequences N1-n are linked to coupling groups which makes it possible to bind to further substances. Said coupling groups include biotin, amino linkers, thiol group or haptens, such as digoxigenin. Using said coupling groups, the nucleic acid sequence N1-n may be

bound to molecules to be labeled specifically and/or may be labeled.

It is possible to label unambiguously the different antibodies with the inventive nucleic acid sequence N1-n by means of the coupling groups generally denoted CG. The binding may be mediated via a protein A/streptavidin fusion protein. This protein binds the constant antibody region and mediates affinity to biotin. Biotinylated nucleic acid sequences bind to such antibodies, thus making selection possible. Figs. 6a - d depict such a selection. Antibodies A, B, C, D and Z have been labeled with the nucleic acid sequences N1, N2, N4, N49 and Nn. The antibodies A, B, C, D and Z are contacted with a matrix to which various antigens A', C', X', D' and Y' have been attached. The antibodies A, C, D, whose antigens A', C' and D' have been attached to the matrix, are bound specifically (Fig. 6b). After removing the antibodies B, Z which have not bound to the matrix by washing, the nucleic acid sequences N1, N4 and N49 may be amplified using labeled primers (Fig. 6c and Fig. 5a, respectively). The amplified nucleic acid sequences N1, N4, N49 are contacted with a detector surface to which the further nucleic acid sequences N1-n containing the detection sequence sections IDP are bound. In this connection, each further nucleic acid sequence N'1-n occupies a particular predefined position on the detector surface (see Fig. 6d).

Figs. 7a - f depict the preparation and identification of a complex code in the form of a flow chart. 50 different nucleic acid sequences are used. The 50 nucleic acid sequences are divided into 5 groups of 10 nucleic acid sequences each (Fig. 7b). A numeral from 0 to 9 is assigned to each nucleic acid sequence of a group (Fig. 7c). One figure of a 5-figure number is assigned to each group of nucleic acid sequences (Fig. 7d).

The code is prepared by taking exactly one nucleic acid sequence from each group of the 5 groups and using said nucleic acid sequence for labeling (Fig. 7e). Each mixture of the 5 nucleic acid sequences taken thus defines a number between 0 and 99999. The numerical value is identified by identification of the 5 nucleic acid sequences (Fig. 7f).

According to another embodiment of the method, the nucleic acid sequences N1-n may be bound to particles P. Binding takes place preferably via activated or activatable groups. These groups include biotin, aminolinkers, thiol groups or haptens such as digoxigenin. Particles P which may be used are polystyrene, silica, polyvinyl chloride, polyethylene, nylon or glass milk particles. It is also possible to make use of particles consisting of virus coats or virus-like particles. The particles P may also be prepared from agents which complex with DNA, such as, for example, polylysine or DNA-binding proteins.

The particles P may be labeled with a variety of nucleic acid sequences N1-n. A particle P may therefore carry a numerical code according to the abovementioned example. Figs. 8a-d depict the preparation of a labeled particle P by way of example. The nucleic acid sequence N1-n carries on one of its ends a biotin group (Fig. 8a). A variety of such biotinylated nucleic acid sequences are dissolved in the same molar ratio (Fig. 8b). The solution is admixed with a predefined amount of particles P which have been coated, for example, with streptavidin (Fig. 8c). A bond between biotin and streptavidin is formed. The nucleic acid sequences N1-n are thus bound to the particle P (Fig. 8d). It is, of course, possible for a plurality of nucleic acid sequences N1-n of the same type to be bound on the particle P. This increases the reliability of the reaction during amplification.

The particles P are preferably from 1 µm to 100 µm in size. They may be fluorescent or may be fluorescently labeled with other agents via a binding reaction. Due to their size and fluorescent property, the particles P may be sorted and isolated by means of a particle sorter. This makes it possible to identify the numberings of individual particles P which are part of a mixture of several particles P.

According to another embodiment of the invention, the particles P may also have coupling groups CG which are suitable for binding to the substance to be labeled (Fig. 9a). These groups include biotin groups, aminolinker groups or thiol groups. They may be attached, for example, at the end of the nucleic acid sequence N1-n. The coupling groups CG are bound to the particle preferably via spacer molecules L (Fig. 9b).

Fig. 9c depicts a particle P with free coupling groups CG. Fig. 9d depicts a particle P in which the substance S is bound to the coupling groups CG.

The substance S1-n may be reacted, for example, with a potential receptor R (Fig. 10a). The receptor R may be labeled with a fluorophoric molecule (Fig. 10b). Particles P carrying a ligand of the receptor R are bound by the receptor R. Particles P containing the receptor R may be separated, owing to their size and fluorescence, from particles P, which have no receptor R bound, by means of a fluorescence-activated particle sorter (Fig. 10c). The bound substance may be identified on the basis of the identification sequence section IDS of the nucleic acid sequence N1-n.

List of reference symbols

N1-n	Nucleic acid sequence
N'1-n	Further nucleic acid sequence
IDS1-n	Identification sequence section
IDP1-n	Detection sequence section
IDS-A, IDS-B	Part section
PBS1	First primer binding sequence section
PBS2	Second primer binding sequence section
CG, A,B,C,D,Z	Coupling groups
L	Spacer molecule
S	Substance
R	Receptor
F11	Fluorophoric group
P	Particle